BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

E5978

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 06-290125
(43)Date of publication of application : 18.10.1994

(51)Int.Cl. G06F 13/00
G06F 12/00

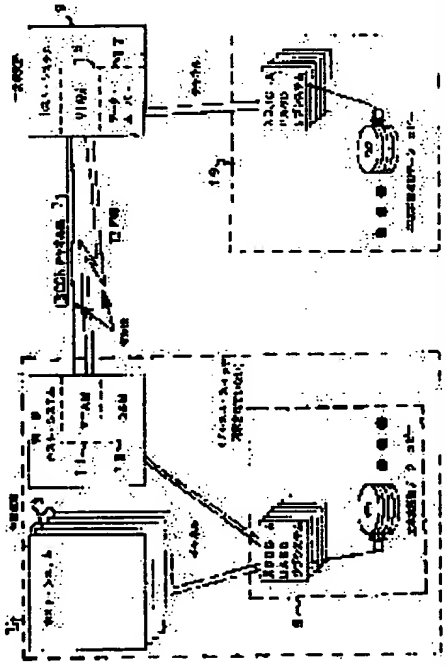
(21)Application number : 06-025562
(22)Date of filing : 23.02.1994

(71)Applicant : INTERNATL BUSINESS MACH CORP <IBM>
(72)Inventor : SHOMLER ROBERT W
MCILVAIN JAMES E

(30)Priority
Priority number : 93 36017 Priority date : 23.03.1993 Priority country : US

(54) METHOD AND SYSTEM FOR REMOTELY BACKING UP AND RESTORING DATA OF MULTIPLEX SYSTEM

(57)Abstract:
PURPOSE: To maintain the consistency of synchronously maintained remote backup copy data in the same state as that for primary-side copy by forming a write token containing a peculiar order number and the address of an external storage device.
CONSTITUTION: In response to each writing operation to a DASD external storage device 5, a write token containing a peculiar order number and the address of the storage device 5 is formed. Then a path is set between a secondary working place 9 and the storage device 5 and the list of newly generated write tokens and the relevant updated data of a token list transmitted with the same or a previous message are transmitted as a message selectively containing each message. At the working place 9, a reserved write queue is maintained for inquiring data with the token in response to each message received from a processor and only the coincident tokens and updated data are scheduled and written in the order of the queue.



LEGAL STATUS

[Date of request for xamination] 23.02.1994

[Date of sending the examiner's decision of r jection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

E5978

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平6－290125

(43)公開日　平成６年(1994)10月18日

(51)Int.Cl.[5]	識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ０６Ｆ　13/00	３５１　Ｍ	7368－5B		
12/00	５３３　Ｊ	8944－5B		

審査請求　有　発明の数５　ＯＬ　（全 13 頁）

(21)出願番号　特願平6－25562

(22)出願日　平成６年(1994)２月23日

(31)優先権主張番号　０３６０１７

(32)優先日　1993年３月23日

(33)優先権主張国　米国（ＵＳ）

(71)出願人　390009531
インターナショナル・ビジネス・マシーンズ・コーポレイション
INTERNATIONAL BUSINESS MASCHINES CORPORATION
アメリカ合衆国10504、ニューヨーク州　アーモンク　（番地なし）

(72)発明者　ロバート・ウェズレイ・ショムラー
アメリカ合衆国カリフォルニア州、モーガン・ヒル、ピードモント・コート　17015番地

(74)代理人　弁理士　頓宮　孝一　（外１名）

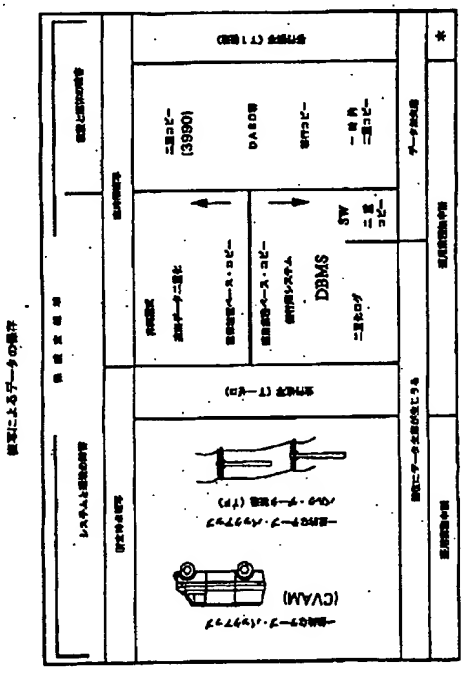
最終頁に続く

(54)【発明の名称】　多重システムの遠隔データバックアップおよび回復を行う方法およびシステム

(57)【要約】

【目的】　本発明は、送信中の更新や、更新コピー発信源が中断された時に受信されなかった更新は除外し、遠隔バックアップコピー・データの保全性を一次側コピーのそれと同じに非同期で維持する方法と手段を提供する。非同期式遠隔データバックアップにより情報処理システムのデータを保存するものであり、そのバックアップは実行中の適用業務を中断せず、さらに伝送中の欠落や主事業所と遠隔事業所間で何らかの中断が発生した時に更新が受信されなかったために生じたデータ欠落が遠隔事業所で解明される。

【構成】主事業所における各書込オペレーションに応答してトークンと固有の順序番号が割り当てられ、このトークンと更新データが遠隔事業所に送信される。この順序番号を用いて順序を確定し欠落した更新データを確認するため順序中にギャップを定める。



1

【特許請求の範囲】

【請求項１】非同期で独立生成された適用業務依存の書込みオペレーションのシーケンスの発生源から遠隔の場所に位置する事業所において、データ・コピー・セットの整合性と回復可能性を与える方法であって、

（ａ）データ発生源における各書込みオペレーションに応答して、固有の順序番号とソース・アドレスを含む書込みトークンを形成するステップと、

（ｂ）データ発生源と遠隔事業所間にパスを確立し、各メッセージが最近生成された書込みトークンのリスト、および同一または以前のメッセージで送信されたトークン・リストの該当更新データを選択的に含むメッセージを前記遠隔事業所に送信するステップと、

（ｃ）データ発生源から受信したそれぞれのメッセージに応答し、データをトークンと照合するために保留書込み待ち行列を遠隔事業所に維持し、適合トークンと更新データについてのみ待ち行列の順で、遠隔事業所で適合トークンと更新データをスケジュールしそして書き込むステップと、を含む方法。

【請求項２】前記方法が更に、

（ｄ）シーケンスの少なくともある部分で生じた中断に応答して、遠隔事業所において維持している保留書込み待ち行列を走査し、データの時間的に整合したイメージを発生源における書込み順序と同じ書込み順序で提示するように、その事業所で受信した順次トークンおよび更新データを解釈するステップ、を含む請求項１の方法。

【請求項３】請求項１のステップ（ｂ）が更に、

データ発生源と複数の遠隔事業所間にパスを確立し、また少なくとも一つのメッセージを１番目の事業所へ送信すること、少なくともいま一つのメッセージを２番目の事業所へ送信すること、あるいは少なくとも一つのメッセージを１番目と２番目の両事業所へ送信することのいずれかを含め、メッセージを送信するステップ、を含む請求項１の方法。

【請求項４】送信中の更新や、更新コピー発生源が中断された時に受信しなかった更新を除き、非同期的に維持されるデータの遠隔バックアップコピーの整合性が一次側のコピーの整合性と同じになるよう保証する方法であって、

コピー発生源と遠隔事業所間にメッセージ・インタフェースを設立するステップと、

発生源における各書込みオペレーションに応答し、発生源アドレスの表示と固有順序番号を表す書込みトークンを生成するステップと、

最近生成されたトークン、書込みトークン、および同一または以前のメッセージで送信されたトークン・リストの該当更新データをリストしたメッセージを発生源から遠隔事業所へ送信するステップと、

遠隔事業所において、

各受信メッセージに応答して、データをトークンと照合

2

するために保留書込み待ち行列を維持するステップと適合するトークンおよび更新データについてのみ待ち行列の順で遠隔事業所において適合トークンおよび更新データをスケジュールしそして書き込むステップと、

中断が発生した場合は、遠隔事業所での更新データ欠落の時点を正確に確立するため、書込み待ち行列と生成メッセージを比較照合するステップと、を含む方法。

【請求項５】プロセッサーと、前記プロセッサーに常駐のオペレーティング・システムと、外部記憶装置と、前記オペレーティング・システムへの呼出しによって前記外部記憶装置へ更新を書き込むため前記プロセッサーで実行している適用業務に応答する手段と、適用業務と非同期的に前記外部記憶装置から遠距離に位置する事業所に前記更新のバックアップコピーを伝播するための手段であって前記更新の前記バックアップコピーを前記遠隔事業所で非同期で書き込むための手段を含むものと、を有するシステムにおいて、

固有の順序番号と外部記憶装置のアドレスを含む書込みトークンを形成するため、外部記憶装置での各書込みオペレーションに応答する手段と、

外部記憶装置と遠隔事業所間に通信パスを設定し、その間でメッセージを送信する手段であって、各メッセージが、新しく生成された書込みトークンおよび同一または以前のメッセージで送信されたトークン・リストの該当更新データを選択的に含むものと、

遠隔事業所にあり、各受信メッセージに応答して、データをトークンと照合するために保留書込み待ち行列を維持し、また合致したトークンおよび更新データについてのみ待ち行列の順で遠隔事業所において、合致したトークンと更新データをスケジュールし書き込むための手段と、

を含む多重システムの遠隔データバックアップおよび回復を行うシステム。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は、非同期式遠隔データバックアップ（遠隔バックアップ複写とも呼ばれる）による情報処理システムのデータ保存に関するものであり、特に主事業所の記憶サブシステムに本拠を置くコピーから遠隔地でデータの複写を行うこと、そしてその複写は、適用業務の実行を中断せず、さらに伝送中の欠落や主事業所と遠隔事業所間で何らかの中断が発生した時に更新が受信されなかったために生じたデータ欠落（曖昧性）の理由を遠隔事業所側で解明できるようなデータ複写に関する。このような中断は典型的に自然災害や人為的災害によって引き起こされ、主事業所での作業とデータを利用不可能にし、作業の継続には遠隔事業所で複写され記憶されているデータと対話することが必須となる。

【０００２】

【従来の技術】データ複写は情報処理システムあるいは計算機システムにおけるデータ保存の一形態である。しかし、データ複写によるデータ保存には多くの要因を考慮に入れなければならない。主事業所での作業やデータが利用不可能になった際に、遠隔事業所で複写され記憶されているデータがそのデータとの継続対話のためのリポジトリーとなることが予測される場合には、このことは特に重要である。複写することについて熟慮すべき要因には、保護定義域（システムや環境の障害、または装置や媒体の障害）と、データ欠落（無欠落、一部欠落）と、他のデータや処理の発生に関連して複写が発生する時間（時点、実時間）と、前記計算機で実行中の適用業務に対する中断の程度と、そして、その複写は適用業務関係のものかあるいは記憶サブシステム関係のものかということが含まれる。前記適用業務関係の複写にはログ・ファイル、データ・ファイル、プログラム・ルーチンを含み、一方前記記憶サブシステム関係の複写には直接アクセス記憶装置（ＤＡＳＤ）アドレスの認識とデータ・セット識別子を含む。

【０００３】次に説明するように、従来の技術には、回復と保存を目的とするデータのバックアップやデータをバックアップするための方法と手段が多く見られる。オリジナル・データ・セットと複写データ・セットとの多少の不整合は許容できるかもしれないが、このような曖昧さを容認できないものにするのは欠落データの発生の看過である。換言すれば、コピー・セットの曖昧さは、回復プロセスがデータ・コピーの状態を判定できないことを意味している。

【０００４】整合性をもつバック・アップ・データ（無欠落複写）を生成するための従来の技術による一方法は、ＤＡＳＤに記憶されているデータの一部を１日１回磁気テープに複写し（一時点複写）、そのテープ媒体をトラックで遠隔地へ輸送する方法である。このように、複写データ・テープのトラック輸送は（１テープ・リール当たり２００メガバイトのデータを記憶しており、このテープ２００リールを主事業所から約８０キロメートル離れた遠隔事業所へ１時間でトラック輸送すると仮定する）平均速度４０、０００ＭＢ／３、６００秒 すなわち１１．１１１メガバイト/秒で通信していることになる。１９５０年代頃のこのようなデータバックアップの実施は適用業務を完全に中断してしまうものであった。さらに、その実施は不便きわまりなく、また主事業所と遠隔事業所との間でデータ状態に１日分の差を生じさせる。

【０００５】さらに他の方法は、バックアップ・コピーを移動するためにＴ１回線のような通信回線を使用する方法である。より現在時間に近い（分単位の時間までも含まれる）最新の遠隔複写が、ＩＭＳ、ＤＢ２のようなデータベース管理システムによって提供される。このような遠隔複写はリアルタイムで行われるが、それは両地点を結ぶ常時使用可能な専用回線を介して伝送されるので、記憶装置関係のコピーよりはむしろ適用業務関係のコピーに依存している。

【０００６】Cheffetz その他による「ネットワークのためのバックアップ・コンピューター・プログラム」に関する米国特許 5,133,065、1992年7月21日付与、は、それぞれのローカル・ノードがバック・アップされるべきローカル・ファイルのリストを作成し、それを送信する宛先となるファイル・サーバーをもつＬＡＮを開示している。このような遠隔生成は、リスト作成およびファイル複写活動をネットワーク・サーバーが開始する場合に生じるトラフィックを減少することを目的とする。議論の余地はあるが、この参照例より以前の技術は中央管理されるファイル選択を教示していた。このことは、結局ローカル・ノードの保安とサーバーの過度使用の妥協をもたらした。これはCheffetz のローカル・ノード生成リストおよびファイル・サーバーへのこのリストの送信によって回避されるものと想定される。

【０００７】Cheffetz が負荷と保安とを調和させることに幾分の懸念のある一時点複写の一形態を述べていることは明らかである。しかしながら、データの整合性に対して、すなわちファイル・サーバーによって実際に複写されたデータの曖昧性を説明することに対しての用意がなされていない。

【０００８】Mikkelsen による「データの無時間差バックアップ複写を行う方法と手段」に関する1991年10月19日出願、審査継続中の米国特許出願第07/781,044号は、論理アドレスと物理アドレスとの一致を形成するための所要時間だけ実行を保留し、その後スケジュールに従い、あるいはその場の都合によって、記憶サブシステムのデータセットを物理的にバックアップすることによって、ＣＰＵの適用業務と並行して、ＤＡＳＤ記憶サブシステムに関し、ＣＰＵでの一時点の整合性をもつ指定されたデータセットをバックアップ複写することを教示している。 Mikkelsen の方法と手段は適用業務を中断せずに選択された記憶装置関係のデータセットの一時点複写を行うのに有利である。それでも、一時点データは同時的複写が開始された時点でのデータの断片として遠隔事業所に到着するという欠点がある。このことは、次に続くデータの更新に比較しデータの状態が何時間も古いものであることを表している。非同期式の遠隔データバックアップは、リアルタイムで行われる複写オペレーションであり、更新は継続的に先送りされるものであることに注意する必要がある。

【０００９】Dishon 他による「同期式磁気ディスク駆動装置における多重コピー・データの仕組み」に関する米国特許 4,862,411、1989年10月29日付与、は、チャネル・コマンド・ワード・チェーンとは無関係に同期している一対のＤＡＳＤへの並列書込みパスを作ることによってバックアップコピーを確実なものにしており、Croc

kett 他による「非同期で稼動する周辺装置の制御」に関する米国特許 4,837,680、1989年6月6日付与、に示されているような単一の制御装置によるバックアップコピー・パスを回避している。ＩＢＭ システム/３７０および同種のシステムのようなシステムにおいては、チャネルと呼ばれる仮想計算機が、パスの設定と、ＣＰＵの主記憶装置と外部記憶装置間のデータ転送とを管理するために使用されていた。制御は、解釈と実行のため外部記憶装置の制御装置へチャネルによって伝えられるチャネル・コマンド・ワード（ＣＣＷ）の形式で与えられていた。

【００１０】Dishon はリアルタイムかつ適用業務無中断の複写を例示しており、それは媒体や装置の障害に備えてデータを保存するが、主事業所側での障害が引き起こすようなシステムや環境の障害は扱わない。

【００１１】Beale 他による「データの冗長コピーを提供するためのデータ記憶システム」に関する米国特許 5,155,845、1992年10月13日付与、は、第一の記憶制御装置によって書込み操作を処理させ、直接リンク（広帯域バス）を介して第二の記憶制御装置に並行通信させることによって２台以上の外部記憶装置へ可変長レコード（ＣＫＤ：カウント・キー・データ方式）バックアップ複写を行う。これは主事業所と遠隔複写場所間のバス長の制限を不要にしている。ＣＫＤの要求/応答アーキテクチャーが長さ（帯域幅）を約150メートル程度に制限している事実から、このような制限が生じている。

【００１２】Beal はDishonの方法で、装置や媒体の障害が発生した場合のみデータの可用性を保つ、リアルタイムかつ適用業務無中断のバックアップ複写を扱っている。

【００１３】別の表現をするなら、非同期でかつ主事業所とは無関係に、遠隔事業所側を更新するために、主事業所でどのようにして書込み更新操作が生成されるかを認識する必要がある。このことについて、主事業所はプロセッサーで並行して実行している一つ以上の適用業務を含んでおり、それぞれの適用業務は「適用業務依存書き込み」と呼ばれるものを生成する。すなわち、記憶サブシステムは、オペレーティング・システムから呼び出されるべき書込み操作や、その待機スケジュールについての知識や認識をもっていない。遅延の変動は局所的なものである。適用業務は単に１台のあるいは同じＤＡＳＤへ書き込むだけではない。事実、初期書き込みとコピー生成のために、待ち行列の長さおよび待ち行列のサービス率がともにＤＡＳＤ記憶装置によって変動するように、適用業務は書込みを異なったパターンに郡別する。これは遠隔事業所で受け取ったコピーが、長時間にわたって順序が揃ったり乱れていたり、また遅延や欠落を起こす確率はかなり高いことを意味している。

【００１４】一言で言えば、非同期でかつ独立して実行している適用業務およびプロセッサーは、ローカル記憶

装置と遠隔事業所に対して書込み操作のストリームを作成する。そのストリームは待機させられ、異なる速さで実行されるため、ほとんどランダムに近いコピー順序をもたらす。

【００１５】

【発明が解決しようとする課題】本発明の目的は、送信中の更新や、更新コピー発生源が中断された時に受信しなかった更新は別として、非同期的に維持される遠隔バックアップコピー・データの整合性を一次側コピーのそれと同一に維持する方法と手段を提供するにある。

本発明の他の目的は、更新コピー発生源と遠隔事業所間のＤＡＳＤ記憶装置の管理レベルで蓄積送信メッセージ・インターフェースを使用するための方法及び手段であって、中断があった場合は遠隔事業所で更新の完全性についての相違や欠落を完全に特定できるような方法と手段を提供するにある。

【００１６】本発明の他の目的は、非同期でかつ独立して生成されたコピーと、主事業所から一ヶ所以上の遠隔データバックアップを行う事業所へ与えられる情報とによって、データ・セットの順序と完全性を回復するための方法と手段を提供するにある。

【００１７】

【課題を解決するための手段】本発明の目的は、以下の構成によるシステムで使用される方法と手段によって満たされる。すなわち少なくとも１台のプロセッサー、少なくとも１インスタンスの前記プロセッサーに常駐のオペレーティング・システム（ＯＳ）、ＤＡＳＤ外部記憶装置、前記ＯＳへの呼び出しを通して外部記憶装置へ非同期でかつ独立して更新を書き込むため前記プロセッサーで実行している適用業務に応答する手段、および前記外部記憶装置から遠く離れた場所において前記更新のバックアップコピーを書込むための手段、から成るシステムである。

【００１８】その方法と手段は、ホスト・プロセッサーでの以下のステップを含む：

（１）外部記憶装置における各書込み操作に応答して、独自の順序番号と外部記憶装置のアドレスを含む書込みトークンを形成すること、（２）遠隔事業所と外部記憶装置との間にパスを設定し、新しく生成された書込みトークンのリスト、および同一または以前のメッセージで送信されたトークン・リストの該当更新データを、それぞれのメッセージが選択的に含むメッセージを送信すること、（３）遠隔事業所において、またプロセッサーから受け取ったそれぞれのメッセージに応答して、データをトークンと照合するために保留書込み待ち行列を維持すること、合致したトークンおよび更新データについてのみ待ち行列の順で遠隔地において、合致したトークンと更新をスクジュールし書き込むこと。

【００１９】本発明においては、プロセッサーのＯＳはＤＡＳＤ外部記憶サブシステムに、ある範囲（トラック

およびシリンダーのセット）のデータ更新シャドウ化が実施されようとしていることを通知する、「シャドウ化」という言葉は遠隔地へコピーを伝達することを意味している。ついでＤＡＳＤ外部記憶サブシステムは書込み活動のためにこの範囲を監視して、書込み操作が進捗していることを順序づけサービスに通知する。この通知は書き込みをされつつあるＤＡＳＤのトラックとレコードを記述するＤＡＳＤトークンによってなされる。順序づけサービスは、ＤＡＳＤが作ったトークンを、協同稼動しているシステムのセット内でシャドウ化されようとするすべての他のオペレーションに関するその書込みオペレーションの時間順を示す順番号と結合させる。この順番号が割り当てられそして書込みデータがサブシステムへ転送された時、サブシステムは入出力書込み操作を開始したホストへ操作完了を知らせる。トークンと順番号は共に、指定されたＤＡＳＤが更新されるという事前情報を提供する非同期メッセージによって、二次事業所へ送信される。

【００２０】続いて、データ転送操作は、順序番号とトークンを与えられた後、サブシステムから更新データを検索し、そして順序番号およびトークンと共にその更新データを遠隔事業所または二次事業所へ送信させる。遠隔事業所または二次事業所ではＤＡＳＤ更新データを受信順に待ち行列に並ばせる。その後遠隔事業所は、主事業所でデータが現れたのと同じ順序でシャドウ（バックアップ）・データ・コピーを更新するため、そのローカル記憶情報のアクセス・オペレーションをスケジュールする。主事業所において災害が発生した場合は、二次データ保管場所における回復プロセスは、遠隔事業所または二次事業所で受信した最後の更新データの時刻におけるそのデータの時間的整合性あるイメージを提出するため二次事業所で受信した順次トークンと更新データを解釈することができる。

【００２１】前記の目的を満足するためには、遠隔事業所または二次事業所での回復プロセスは、主システム・コンプレックスが、装置の破損はないが、完全に全体的な故障中断を被ったときに主事業所においてＤＡＳＤに記憶されていた筈のデータと同等なＤＡＳＤデータのイメージを回復場所において提供できることが必要である。遠隔事業所における回復プロセスの実行に続いて、主事業所において幾分早期に行われた初期システム始動IPLに続いて現れていた筈のデータと全く同様な二次データがシステム・プロセッサーで実行される適用業務に現れる筈である。この「早期に」という時間量は、主事業所から回復事業所への距離、主事業所と回復事業所間で利用できる転送チャネルの帯域幅、およびデータ・ムーバーのために利用できる処理能力等の相関要素で変わりうる「データ欠落のウインドウ」である。

【００２２】

【実施例】本発明は、システムに含まれる各中央演算処理装置（ＣＰＵ）が、ＩＢＭ社のＭＶＳオペレーティング・システムを持つＩＢＭ社のシステム/３６０またはシステム/３７０のアーキテクチャーを採用したＣＰＵで構成されるシステムにおいて都合良く実施できる。ＩＢＭ社のシステム/３６０アーキテクチャーを採用したＣＰＵに関しては、Amdahl　その他による「データ処理システム」に関する米国特許 3,400,371、1968年9月3日付与、に詳述されている。外部記憶装置を共用アクセスする複数ＣＰＵを含むシステム構成に関しては、Luiz その他による「複数ＣＰＵと共用装置アクセス・システムにおける経路非依存の装置予約と再接続」に関する米国特許 4,207,609、1980年6月10日付与、に記述されている。Luiz は仮想コンピューターすなわち「チャネル」管理と、チャネルから発せられ記憶サブシステムにより受信され解釈される順次または連鎖のチャネル・コマンド・ワード（ＣＣＷ）および同様なコマンドによる外部記憶装置の制御についても記述している。

【００２３】ＭＶＳオペレーティング・システムもＩＢＭ社の出版物、GC28-1150、「ＭＶＳ/拡張アーキテクチャー・システム・プログラミング・ライブラリー」、ボリューム１に記述されている。ローカル・ロック管理、割込みまたは監視によるサブシステム呼出し、タスクの通知と待機などの、標準ＭＶＳの詳細およびその他のオペレーティング・システム・サービスの詳細は省略されている。これらのＯＳサービスは当業者に十分に知られているものと信じられる。

【００２４】ここで、図１を参照するに、システム、装置のいずれかの障害を起こしがちなシステムにおいて、複写によりデータを保存することにかかわる多様な要因を概念化して説明されている。保護定義域は図示しているようにシステムと環境を包含し、装置や媒体のレベルでの可用性の改善に限定される。

【００２５】図１に示されているように、一時点複写は歴史的にみて適用業務を中断するものであった。それはバックアップの複写オペレーションが完了するまで実行を延期することを必要とした。現在のシステムはＤＡＳＤ記憶サブシステムからテープへのスケジュールされた転送を行う。一方、先に言及したように、Mikkelsenの審査継続中の出願は適用業務実行の中断を排除し、一時点バックアップをその場の都合に合わせてスケジュールすることを可能にするのではあるが、バックアップする事業所または貯蔵所におけるそのデータの状態は、主事業所における現在のデータ状態より常に遅れている。

【００２６】図１に示されているように、通常、リアルタイム複写は、適用業務の実行を中断せず、データの欠落が最小限または皆無であり、そして遠隔事業所を主事業所と同じ最新状態に維持する。ＤＡＳＤ群を含み、バックアップ複写することからデータ・ブロックの特別に区分された冗長コード化グループを装置へ非同期式で書込むことにまで及ぶ、前記リアルタイム複写方法は、装

置と媒体の障害に関してはデータの可用性を拡張するだけであり、主事業所が利用不可能になった場合を考慮していない。

【００２７】遠隔事業所への通信に基づく他のリアルタイム複写の方法は、専用バスを必要とし、プロセッサー、記憶サブシステム、および遠隔事業所を完全に閉鎖することになる可能性が高い。これの典型的なものはリレーショナル・データ・ベース、自動化銀行システム等のログ・ベースのトランザクション管理システムである。このようなシステムでは、適用業務だけが、コピーの起源と、障害後のデータ状態をトランザクションと一致した状態に回復すべきことの必要性を認識している。

【００２８】次に図２を参照するに、複数台のＤＡＳＤサブシステム５に共用アクセスする１台以上のシステム３と遠隔または二次プロセッサー設置事業所９を連結するインターフェース７で構成される主事業所１が図示されている。本発明のために、これらのシステムの１台は「選定主システム」に指定される。選定主システムはオペレーティング・システム（ＯＳ）をもつプロセッサーを含んでいる。また、ＯＳはデータ直列化ムーバー（ＤＳＭ）１３と仮想記憶通信アクセス方式（ＶＴＡＭ）１１、またはインタフェース７と遠隔事業所９間の通信を達成させるための同様なものを含む。

【００２９】ＤＳＭの直列化機能部分は各書込み作業に書込み順トークンを割り当て、ＶＴＡＭ１１が二次事業所（遠隔事業所９）の受信側システム（ＶＴＡＭ１５およびデータ・ムーバー17）へ送信するためのメッセージにこれらのトークンを入れる。また、ＤＳＭ13のデータ・ムーバー機能部分は主ＤＡＳＤに書込まれた変更済みデータ・レコードを受け取り、それらのデータ・レコードとトークンをＶＴＡＭ１１が二次事業所９へ送信するためのメッセージにの形に形成する。ＶＴＡＭ１１はシステム・ネットワーク体系（ＳＮＡ）のホスト・プロセッサー部分である。ＳＮＡはＩＢＭシステム・ジャーナル、1979年18巻2号に掲載されている P.E.Green 、「ネットワーク体系とプロトコル入門」に記述されている。ＶＴＡＭのオペレーションと実施に関する追加の詳細については、Harris 他による「透過サービス・アクセス機構の適用業務を遠隔ソースへ相互接続するための装置と方法」に関する米国特許 4,914,619、1990年4月3日付与、、および Astride 他による「エミュレートした入出力制御装置および共通コマンド変換テーブルと装置アドレス・テーブルを包含する複数プロトコル入出力通信制御装置」に関する米国特許 4,855,905、1989年8月8日付与、を参照することができる。ＶＴＡＭ１１とＶＴＡＭ１５は機能的には主事業所１と二次事業所９のいずれでも作動できることに留意すべきである。

【００３０】さらに図２を参照するに、ＶＴＡＭ１１、ＶＴＡＭ１５、および主事業所１と二次システム９とのＳＮＡまたはESCON接続７の目的は、その間でメッセージの通信をすることである。これらのメッセージは、シャドウ化（バックアップ）データ・エクステントの確立、進行中の更新を表示するトークン、および二次側データ・コピーに適用される変更済みデータをもつトークンを伝送する。ＶＴＡＭは主システムと二次システムを結合する単一または複数のチャネル間接続の広帯域通信ネットワーク上で作動できる。Ｔ１回線、Ｔ３回線、またはその他の通信サービスは、ESCONや他のチャネル間直接接続がサポートしうる距離よりも長距離をサポートするために使用される。二次システム9は主事業所１から任意の距離であってよい。

【００３１】同様に、二次システム９のプロセッサー部分に常駐のＯＳは、上に述べたＶＴＡＭ１５のＯＳベースのサービスに加えてデータ・ムーバーの構成要素１７をもっている。二次システムのデータ・ムーバー１７は、ＶＴＡＭ１５を介して主システムから送信されるメッセージを受信し、順序付けトークンを抽出、維持し、そして更新データを二次事業所にあるＤＡＳＤコピーに適用する。

【００３２】バックアップオペレーションは、Mikkelsen の審査継続中の出願に記述されているようなＴ-ゼロ（Ｔ０）・コピー・セッションを確立するための方法と同様に主システムによって、選定されたＤＡＳＤ、ボリュームまたはエクステントに対して確立される。Mikkelsen の方法によれば、主システムのデータ・ムーバーは識別されたデータ・エクステントのＴ０所定時コピーを読取り、そしてバックアップデータの最初の二次コピーを確立するため、そのデータを二次側へ送信する。ひとたびバックアップセッションが開始されると（タイム・ゼロ・コピーの付随の有無にかかわらず）、ＤＡＳＤサブシステムは、データの書き込みのためそれらのエクステントを監視し、書込みオペレーションが開始されると以下の作業を行う。

【００３３】主事業所の外部記憶サブシステムにおける書込みオペレーション

次に図３を参照するに、図２と同じ構成において、バスに付した番号による順次作業を表示している。このことについては、主事業所ではプロセッサーとＯＳのチャネル部分は、バックアップＤＡＳＤアドレスへのデータ書込み順序を開始するＣＣＷを生成する。これは主事業所ＤＡＳＤサブシステム５（ＥＣＫＤ位置指定レコードＣＣＷに関しては、これはその順序に対する数個のＣＣＷの最初のＣＣＷであろう）によって処理され、そしてデータは、書き込みをするシステムからバス（１）を通って主事業所のＤＡＳＤへ転送される。主事業所のＤＡＳＤサブシステムはバス（２）を通ってＤＳＭの直列化機構１３（それはＤＡＳＤへデータを書込んでいるシステムと同じシステムに含まれる機構とは限らない）へ書込みトークンを送信する。要求トークンはバックアップセッション、装置、書込みトラックとレコードを指定す

る。また要求トークンは固有のサブシステム事象トークンをもつことができる。ＤＳＭの直列化機構１３は、そのＤＡＳＤトークンに大域順序番号を割り当て、順序とＤＡＳＤとの複合トークンを次のメッセージへ付加する。そのメッセージは、二次事業所または遠隔事業所９へ送信するためＶＴＡＭ１１へ渡され、バス（４）を通って事業所９へ送信される。

【００３４】ＤＡＳＤデータ書込みを実行しているＤＡＳＤシステム５は、そのデータ書込みオペレーションを完了すると、バス（３）を通って要求したシステムへ書込み完了の信号を送る。（もしこれがホスト・チャネル・プログラムの終了であれば、書込み完了は要求プログラムへ通知され、そうでなければチャネルは、次のオペレーションでチャネル・プログラムを継続するものと、その信号を解釈する。）最初の書込みオペレーションに非同期の別個のオペレーションとして、またバス（５）を通って記述されるＤＡＳＤ供給トークンからのＤＡＳＤアドレス情報を用いて、ブロック１３に常駐のデータ・ムーバーは、ＤＡＳＤサブシステム５からの変更データを読取る。ついで、データ・ムーバー１３は、そのデータをその順序トークンと共に、次のメッセージへ付加し、メッセージは、ＶＴＡＭ１１、インターフェース、ＶＴＡＭ１５、データ・ムーバー１７、主事業所・二次事業所間のバス（６）を包含する回路を通って二次事業所へ送信される。（書込み通信量をこなすため複数のデータ・ムーバー・プロセスがあってもよい。）

肯定リターン・トークン

本発明のより強力な実施例は、ＤＳＭ１３が割り当てられた大域順序番号と共に書き込みトークンをＤＡＳＤサブシステム５へ返送することを伴う。続いてＤＡＳＤサブシステムは、変更データがデータ・ムーバーへ送信される時に、ＤＳＭ供給トークンを変更データに付加する。これは同一のレコードに急速に複数の更新がなされる場合の潜在的不確実性を排除する。このことは前記の基本設計がプログラミングによって適合しなければならない条件である。この肯定リターン・トークンによる方法は、ＤＡＳＤサブシステムが保留トークンと付随するデータ（ポインターによる）の持続的な交互の入出力オペレーションの作業リストを維持することを必要とする。これはある種のトラック・キャッシュＤＡＳＤサブシステム（ＤＡＳＤレコードによる書込み更新バックアップ）に適合しない設計である。

【００３５】二次事業所におけるデータ複写オペレーション

バックアップセッションがひとたび確立されると、二次事業所９は、保留書込み通知と、バックアップコピーに保存しているＤＡＳＤ１９への更新された書込みデータを受信する。ＶＴＡＭ１５は主事業所１からのメッセージを受信し、それらを二次事業所のデータ・ムーバー１７へ渡す。これらのメッセージはそれぞれ、主事業所側データ・ムーバーによって作られ二次側データ・ムーバーにより解釈される、三つのコンテント・セグメントから成り立っている。

【００３６】次に図４を参照するに、各メッセージの三つのセグメントが図示されている。それらはＭ０、Ｍ１、およびＭ２として示されている。

【００３７】Ｍ０はデータ・ムーバー１３からデータ・ムーバー１７への見出しで、メッセージ・ストリームの論理連続性を維持するのに役立ち、またメッセージ・コンテントを識別する。メッセージ・コンテントは、トークン・データ伝送（Ｍ１-Ｍ２）を含み、バックアップセッションを確立または終了し、一時点複写の初期コピーまたは他の形式の初期データ・コピーを伝送し、二次側から主事業所側へ論理肯定応答（ACK）および例外通知と回復処置のため通信することができる・Ｍ１は、先行するメッセージの後で、一次側ＤＳＭによって割り当てられたトークンのセットを含んでいる。これは二次側にデータが存在しない場合の書込オペレーションが進行中であることを表す。

【００３８】Ｍ２はフィールドのセットを含み、それぞれのフィールドはトークンとそれに付随する書込み（更新）データを含む。

【００３９】メッセージの長さと伝送の頻度は、システム間伝送の能率と、進行中の書込みオペレーションについて二次側への通知に対する遅延時間とのバランスで決まる設計考慮事項である。複数の保留トークンとデータとをメッセージとして一括し、伝送のオーバヘッドを多数のＭ１およびＭ２メッセージ・エレメントに振り分けることにより、能率は向上される。より短い、そしてより頻繁に送信されるメッセージは、より大きな合計伝送オーバヘッドを生じることになるが、二次側を一次側とより近く同期（遅れが小さい）した状態に保つ働きがある。この概念は、メッセージ・バッファーが満杯になった時か、あるいは先行するメッセージが送信された後の時間間隔が満了した時かのいずれか早い方で、一次側データ・ムーバーはメッセージをＶＴＡＭへ渡すということにある。メッセージ・バッファーのサイズおよび時間間隔は共に選択できるものとする。

【００４０】各トークンは、二次側データ・ムーバーがどの物理ＤＡＳＤレコードが更新されるかを判別し、またこれらの書き込みを順序トークンが指定されたのと同じ順序（主事業所側ＤＡＳＤへの書込みオペレーションを行った適用業務から見た順序と同じである）で配列するのに十分な情報を含んでいる。二次側データ・ムーバーは、トークンがＭ１セグメントに受信されるときに、まず最初に各トークンを見る。データ・ムーバーは、これらのトークンを保留書込み待ち行列を維持するために使用する。各トークンに対するデータがＭ２セグメントに受信されるとき、データは保留中の書込み待ち行列に含まれているそのデータに対するトークンと照合される

（図３および図４に示されているように）。

【００４１】二次側データ・ムーバー１７は、二次事業所ＤＡＳＤ記憶装置１９への書込みオペレーションを、保留中の書込み待ち行列入力の順でスケジュールする。これはバス（７）を通る待機転送として図３に示されている。言い換えるならば、先行する待ち行列エレメントがＤＡＳＤへの書込みをスケジュールし終えるまでは、所与の待ち行列記入事項に関するデータは書込みをスケジュールされない。データは、それが主事業所側データ・ムーバーによって与えられた時点の関数としてＭ２メッセージ・セグメントへ到着する。これらのデータの到着は、厳密な更新順にはならないだろう。かくして、活動の任意の時点に、保留中の書込み待ち行列は、完全な記入事項のトークンと付属の書込みデータ、書込みデータがない不完全トークン、これに続くさらにより多くの完全トークンと不完全トークンという一連のシーケンスをもつことができる。このことは図５に図示されている。

【００４２】主事業所側の障害時の二次側でのＤＡＳＤデータの回復：上述のメッセージは、バックアップされる主事業所側のデータの回復のため二次側システム９において、情報を提供する。もし主事業所１が、二次側システムでのデータ回復が求められるような災害を被ったら、システムのオペレーション・プロシージャーは、バックアップされるデータの全部または一部のために二次側における回復オペレーションを呼び込むことができる。

【００４３】次に図５を参照するに、第二事業所を使用するシステムに対する回復プロセスは、保留中の書込み待ち行列から、待ち行列中の最初の不完全トークンまでの保留中のデータ書込みを完了することによって、開始される。これに関連して言えば、データは主事業所におけるいくらか早い時点、すなわち主事業所の障害よりもいくらか以前の時点の所まで整合性を持っている。これは図５の「Ａ」で示されている。このデータは障害時点において、図５の「Ａ」「Ｂ」間の記入事項によって表される保留中の書込み待ち行列中の記入事項の残余分と、そして主事業所側でスケジュールされていて実行されていただろうが、それに対してＤＳＭが割り当てた順序トークンが、二次側で図５の後続位置「Ｂ」に示されているＭ１セグメントに受信されなかった書込みオペレーション分だけ、主事業所側より遅れることになる。

【００４４】もしシステム回復が一時点まで整合性のあるデータを使用できるならば、それは直ちに回復することができる。これは、主事業所側でシステムのハードウェアやＤＡＳＤのデータに損壊のない故障が発生した場合（すなわち進行中のジョブと非持久記憶機構のデータは失われてしまい、オープン・データ・セットは整合性のない状態になるだろうが、ＤＡＳＤサブシステムに書き込まれたデータはアクセス可能である）、主事業所側でのシステムの電源立ち上げおよび初期プログラム・ロード（ＩＰＬ）のプロセスとほとんど同じであると期待される。

【００４５】オペレーションの回復・再開始に、主事業所側で何のデータが欠落したかに関してより多くの情報を必要とするような適用業務に対しては、障害前に主事業所側で進行中の更新が入っていたＤＡＳＤのトラックおよびレコード、ならびにオペレーションの順序（それが回復に際して必要ならば）を識別するために、保留書込み待ち行列コンテントが解釈され得る。一部の更新に対しては、先行するいくつかの更新が受信されていなかったため、これらの更新が遠隔ＤＡＳＤに反映されていないにもかかわらず、更新されたデータでさえも遠隔事業所で問い合わせができる。これらの更新は図５の「Ａ」「Ｂ」間の順番号１３１および１３５として示されている。

【００４６】さらに、バックアップセッションを確立することの一部として送信された情報、あるいは協同作動している記憶装置管理システム間で別個に送信された情報は、保留書込みのＤＡＳＤアドレスをデータセット（保留書込はこれの一部である）に関連づけるために用いられる。識別されていなかったことになる唯一の進行中の書込みは、主事業所側で作成されたが送信されなかったＭ１中の書込みおよび障害発生時に伝送中であったＭ１メッセージである。従って、データセットを整合性ある状態に復元し、システム・オペレーションを再開するために、必要ならば、回復プロセスは、正確にどのデータが欠落したかを障害発生に極めて近い時点まで識別することができる。

【００４７】この新方法は、必要なすべてのＤＡＳＤにわたる更新順序の保存を可能にし、そして主事業所側の書込み性能への影響を最小化する（ＤＡＳＤ入出力書込みオペレーションの応答時間の増加を最小化する）記憶装置ベースのデータの遠距離バックアップを行うための手段を定義する。

【００４８】従前の既知の方法は次のいずれかであった；（１）データ・ベース・トランザクション・コミット点の認識のような、記憶装置ベースのデータのもつ知識以上のものを要求する、（２）データの整合性を保ったまま、バックアップされたすべてのＤＡＳＤデータにわたるデータ更新順序の保存ができない、（３）ＤＡＳＤ入出力書込みオペレーションと同期して作動するので、入出力オペレーションの応答時間を容認できない程度にまで増加する。

【００４９】拡張

本発明の方法と手段に対する一つの拡張は、その組み合わせに関する優れた柔軟性を利用することである。実例として、バックアップセッションは複数の二次事業所で確立することができ、何らかの主事業所側データを一ヶ所の二次事業所へ、また何かを他の事業所へ、そして何

15

かを二ヶ所以上の事業所へ送信することができる。

【００５０】請求項に列記されている精神と範囲からから逸脱することなしに、本発明のその他のあらゆる拡張をすることが可能である。

【００５１】

【発明の効果】本発明によれば、送信中の更新や、更新コピー発生源が中断された時に受信しなかった更新は別として、非同期的に維持される遠隔バックアップコピー・データの整合性を一次側コピーのそれと同一に維持することができる。また本発明によれば、中断があった場合は遠隔事業所で更新の完全性についての相違や欠落を完全に特定できるような方法と手段が提供される。また、非同期でかつ独立して生成されたコピーと、主事業所から一ヶ所以上の遠隔データバックアップを行う事業所へ与えられる情報とによって、データ・セットの順序と完全性が回復される。

16

＊【図面の簡単な説明】

【図１】複写によってデータを保存する際に考慮すべき多様な要因を示す概念図。

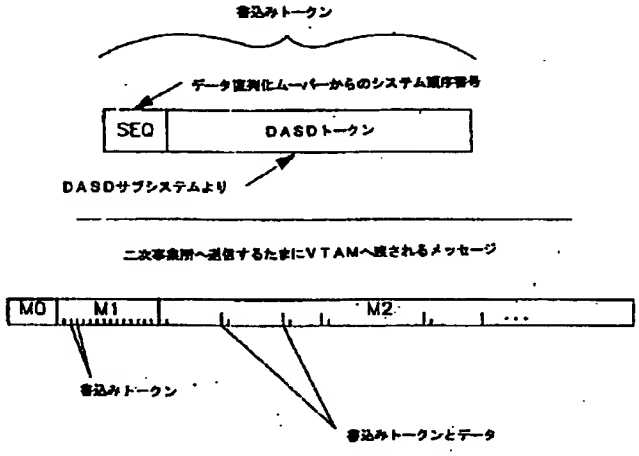
【図２】ホスト・プロセッサー、ホストに接続された外部記憶装置、データがバックアップされる遠隔事業所、および外部記憶サブシステム/遠隔事業所通信インターフェース相互間におけるネットワークの構成図。

【図３】活動順に番号付けをしたパスを表示した図２と同じ構成のネットワーク構成図。

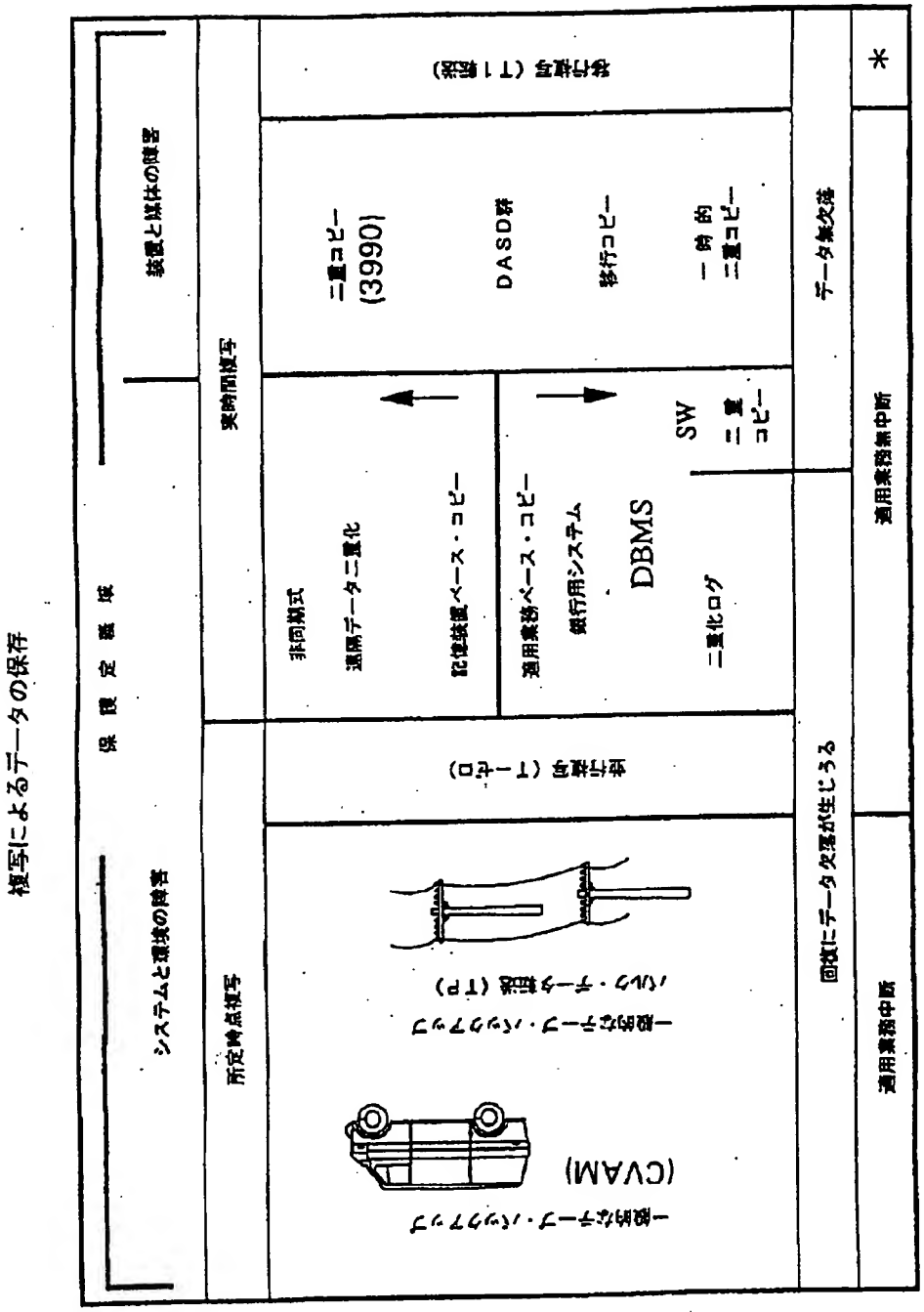
【図４】トークンのフォーマット、順番号、およびサブシステムと遠隔事業所間で使用され通信されるメッセージのフォーマットを示す図。

【図５】主事業所における直前の書込み更新活動のスナップショットとしての遠隔事業所におけるメッセージとトークンの待ち行列を示す図。

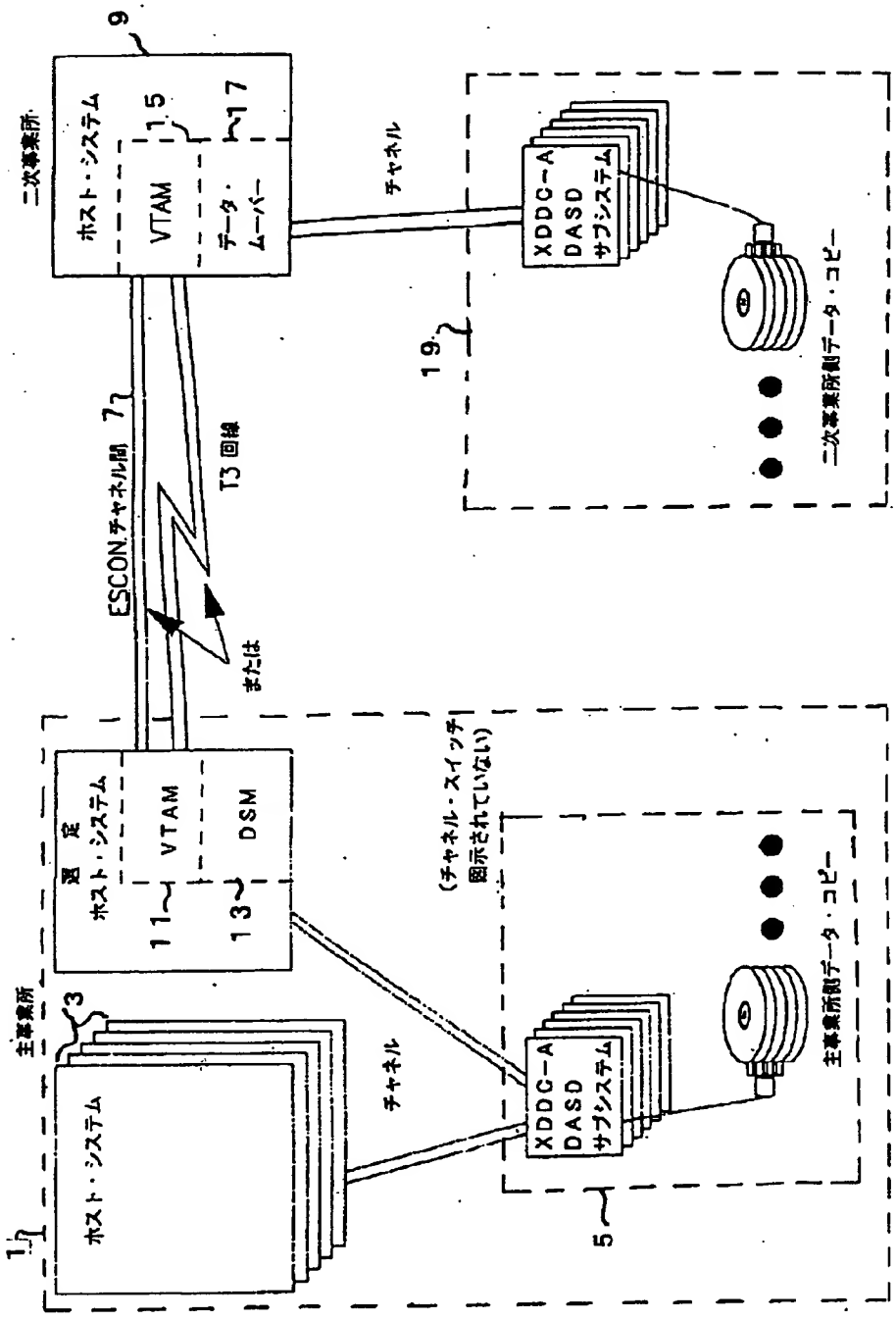
【図４】



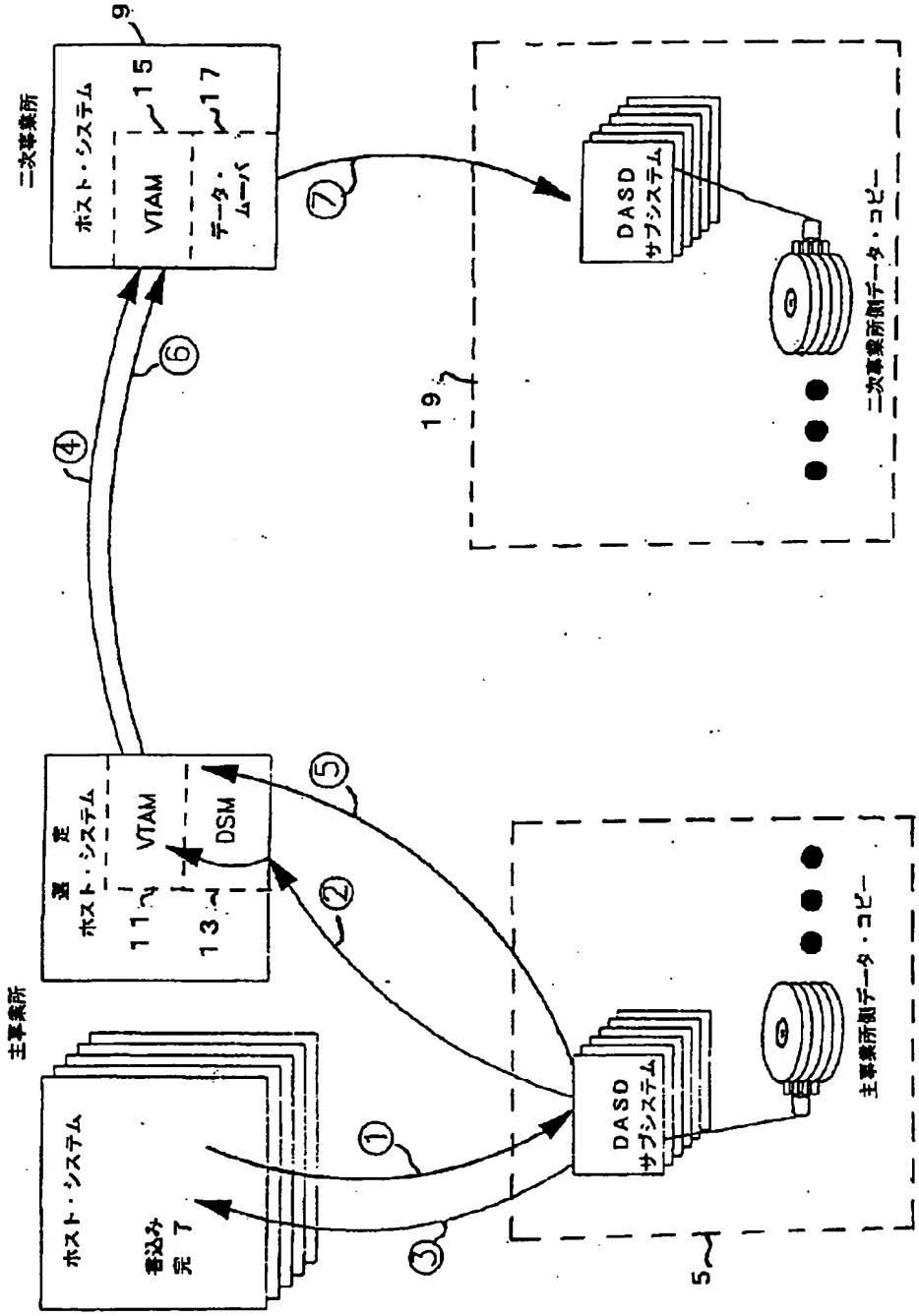
【図1】


複写によるデータの保存
保護定義域
システムと環境の障害
装置と媒体の障害
所定時点複写
実時間複写
一般的なケース・シナリオ
(CVAM)
一般的なケース・シナリオ
ハング・アップ状態 (TP)
非同期式
遠隔データ二重化
記憶装置ベース・コピー
DBMS
二重化ログ
適用業務ベース・コピー
銀行用システム
SW
二重コピー
二重コピー
(3990)
DASD群
銀行コピー
一時的
二重コピー
移行複写 (テープ転送)
非行複写 (テープ)
データ集欠落
回復にデータ欠落が生じうる
適用業務中断
適用業務集中断
*

【図2】

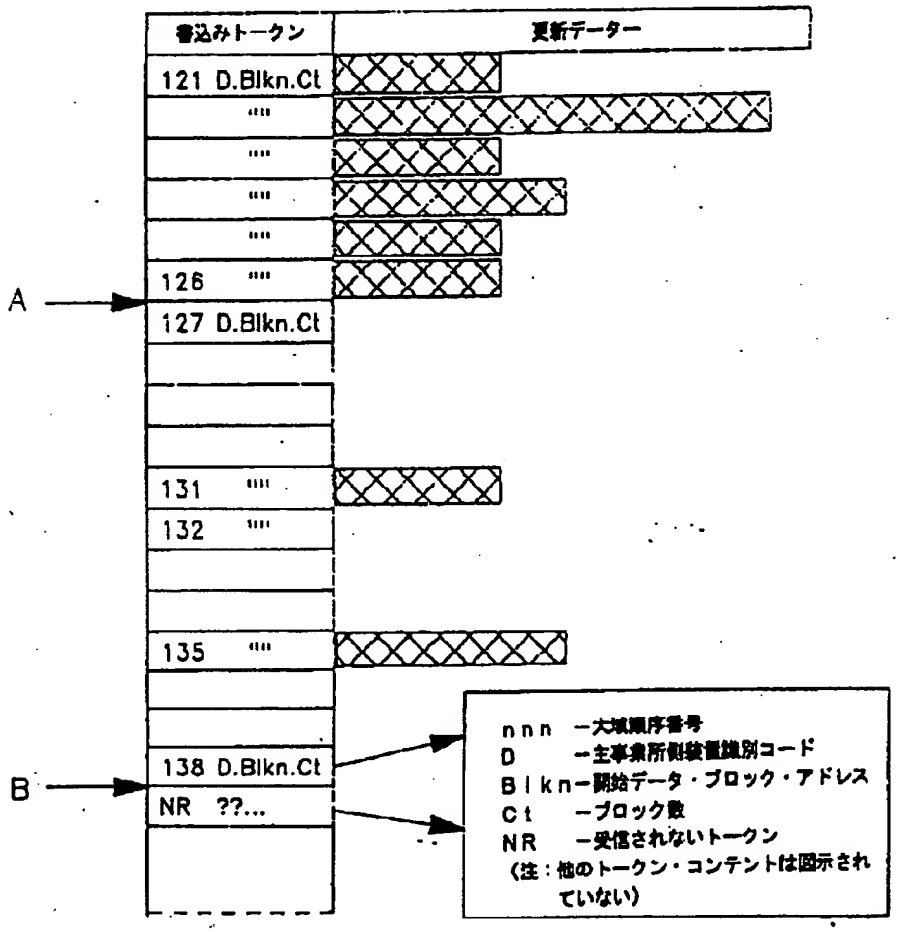

9
二次事業所
ホスト・システム
VTAM
15
データ・ムーバー
17
チャネル
XDDC-A
DASD
サブシステム
19
二次事業所データ・コピー
ESCONチャネル間
7
T3回線
または
選定
ホスト・システム
VTAM
11
DSM
13
(チャネル・スイッチ
図示されていない)
主事業所
3
ホスト・システム
チャネル
XDDC-A
DASD
サブシステム
主事業所データ・コピー
1
5

【図3】


二次事業所
ホスト・システム
VTAM
データ・ムーバ
15
17
9
DASD
サブシステム
二次事業所側データ・コピー
19
主事業所
ホスト・システム
書込み完了
DASD
サブシステム
主事業所側データ・コピー
5
VTAM
DSM
11
13
①
②
③
④
⑤
⑥
⑦

【図5】

二次事業所側における保留書込み待ち行列



フロントページの続き

(72)発明者　ジェームズ・イー・マッキルベイン
アメリカ合衆国カリフォルニア州、サン・
ホゼ、シルバ・オーク・コート　1118番地